REMARKS

Claims 1-5, 7-14, 16 and 19-30, of which Claims 1, 10, 19, 20, 21 and 26 are the independent claims, are pending in the application.

Claims 1-5, 7-9, 19 and 21-25 have been rejected under 35 U.S.C. §101 as directed to non-statutory subject matter.

Claims 1-5, 7-14, 16 and 19-30 have been rejected under 35 U.S.C. §112, second paragraph, for failing to particularly point out and distinctly claim the subject matter that the applicant regards as the invention.

Applicants respond as follows.

Claim Amendments

Support for the amendments to Claims 1, 10, 19 and 20 is found at least on Specification page 7, lines 13-16 and lines 26-29 and page 10, lines 3-8 and Fig. 2 as originally filed.

New Claims 34-36 have been added but merely repeat unentered Claims 31-33. Support for Claims 34-36 is found on pages 11-13 of the specification as filed and, in particular, page 11, lines 16-19, page 12, lines 22-24, page 12, line 25 to page 13, line 6. Additional support for Claims 34-36 is found on page 7, lines 13-16 and page 10, lines 18-25 of the specification, as filed.

No new matter is introduced by this Amendment.

Applicant's Invention

The present invention is a method and an apparatus for analyzing a sequence from a subject's genome by assigning this sequence to a class. The method operates by accepting two inputs: (1) a set of "known biological fragments", *i.e.* characteristic biological sequences obtainable from available databases, *e.g.* protein domains; and (2) a subject genome sequence to be analyzed.

During the operation of the method of the invention, a "feature vector" is created for each genome sequence that is being analyzed. The "feature vector" is a list of numerical scores of each fragment from the set of "known biological fragments". The numerical scores are obtained by measuring the likelihood of finding a given fragment within the subject genome sequence that

is being analyzed. Advantageously, the use of these "feature vectors" allows one to compare genome sequences of variable length since each "feature vector" always has the same number of elements equal to the number of fragments in the set of "known biological fragments". This feature of the method is referred to as "uniform" representation.

During the operation of the method of the present invention, the "feature vector" represents the genome sequence being analyzed in an a manner that is readily accepted by a machine (a classifier, indexer, clusterer or, generally, a genome analyzer). Implicit in any process of assigning an object to a class is comparison between the object being assigned and the objects that had already been assigned to this class. In the context of the present invention, the genome sequence being analyzed is referred to as a "query" (specifications page 11, line 5), while the sequences that had already been classified are referred to as a "training set" (specification page 10, lines 18-20).

One skilled in the art will immediately appreciate that when more than one subject's genome sequences are being analyzed, the earlier analyzed sequences can be considered a "training set" with respect to later analyzed sequences. However, in other embodiments of the present invention, specially designated "training sets" of sequences can be provided to create classes that will be used in classifying a subject's genome sequence (specification page 10, lines 18-25 and Example, pages 11-13).

Although methods for creation of classes are not a part of the inventive concept, they are described in the instant application. When the method of the present invention is applied to a "training set" of sequences, a set of training feature vectors is produced. Any of a number of classification algorithms can be applied to such a "training set" of feature vectors. In one embodiment, "N12" indexing technique is applied (specification page 11, lines 1-12). Other techniques include k-mean clustering, nearest neighbor analysis or support vector machine technique, all described in the specification on pages 10-11. Still other embodiments of the invention can create classes by employing principal component analysis (PCA) (specification page 10, lines 1-2).

Whichever classification scheme is used, a classification of sequences is produced, *i.e.* classes of sequences are defined. An implicit purpose of any classification system, such as the sequence-analyzing system of the present invention, is to answer a question: "Does an object O

belong to a class C?" In the instant invention, once the classes of sequences are defined, the question is "Does the subject's genome sequence S belong to a sequence class C?" The answer to this question, a "yes" or "no", can be considered an elementary output of the method and the apparatus of the instant invention.

Rejection of Claims 1 - 5, 7 - 9 and 19 under 35 U.S.C. §101

Claims 1 - 5, 7 - 9 and 19 have been rejected under 35 U.S.C. §101 as being directed to non-statutory subject matter. The Office Action states that the claims are drawn to methods of manipulating data which do not produce a concrete, tangible and useful result. The Office Action further states that examples of methods that met statutory requirements of 35 U.S.C. §101 include results such as a "dollar value", "buy" or "sell" recommendation or "yes" or "no" answer. The Office Action further states that the value obtained [by data manipulation] must be immediately useful, without further transformation or manipulation by the user.

While the Applicant believes that examples of the methods that satisfy 35 U.S.C. §101 are correct, the Applicant submits that the summary of the law given in the Office Action is imprecise.

M.P.E.P. § 2106(II)(A) states that

> "[o]nly when the claim is devoid of any limitation to a practical application in the technological arts should it be rejected under 35 U.S.C. §101."

For example, in AT&T Corp. v. Excel Communications, Inc., 172 F.3d 1352, 50 USPQ2d 1447 (Fed. Cir. 1999) claims were drawn to a long-distance telephone billing process containing mathematical algorithms that included determination of whether a particular switch was ON or OFF. The claims were held to be directed to patentable subject matter because "the claimed process applies the Boolean principle to produce a useful, concrete, tangible result without pre-empting other uses of the mathematical principle." AT&T Corp. v. Excel Communications, Inc., 172 F.3d 1352, 1358, 50 USPQ2d 1447, 1452 (Fed. Cir. 1999).

As submitted above by the Applicant, the method of the present invention produces a "yes" or "no" answer to a question: "Does the subject genome sequence S belong to a sequence class C?", which is similar to the question of the AT&T Corp. "Is the PIC switch ON or OFF?".

The Applicant submits that independent Claims 1 and 19 are indeed drawn to methods that, as their last steps of "analyzing" the subject genome or protein sequence, answer the above question, thus, in effect, producing a "yes" or "no" type answer. Even more explicitly, Claim 19 recites "classifying the subject protein sequence into a ... class". The Applicant submits that it would be extremely difficult to assign an object to a class without answering "yes" or "no" to a question of whether *this* object belongs to *this* class.

Further, like in AT&T Corp., where the method of determination of whether a switch was in an ON or OFF position was *further used* for billing purposes, the assignment of a subject's genome sequence to a known class of sequences produces knowledge of structure and/or function of a gene or a gene product (*e.g.* a protein or an RNA) and is used in medical diagnosis, drug design, and genome annotation. (See, for example, pages 1 and 2 of the specification).

To further emphasize the foregoing patentable features of the method of the present invention method, new Claims 34-36 are presented. New Claim 34 is based on Claim 21 amended to recite the steps that can, in one embodiment, be used to create sequence classification. New Claim 35 is based on Claim 19 and recites the steps that can, in one embodiment, be used to create sequence classification. New Claim 36 is made dependent on Claim 26 (an apparatus claim) and recites the steps that can, in one embodiment, be used to create sequence classification.

Thus, the claimed invention as a whole accomplishes the requisite practical application in the technological arts by classifying (indexing, clustering or, generally, analyzing) genome sequence. Such classification (or, generally, analysis) possesses a level of "real life value" in keeping with M.P.E.P. § 2106(II)(A) and 35 U.S.C. §101.

The Applicant proposes the foregoing and asks for acceptance by the Examiner in order to move prosecution of the subject application forward in good faith. The claims as presented and proposed recite the present invention in terms that are drawn to a patentable subject matter, novel and not made obvious by the cited art.

It is respectfully requested that the §101 rejection be withdrawn.

Rejection of Claims 1 - 5, 7 -14 and 16 - 20 Under 35 U.S.C. §112

Claims 1 - 5, 7 -14 and 16 and 19-30 have been rejected under 35 U.S.C. §112, second paragraph as indefinite.

The Office Action states that the "metes and bounds" of the limitation "including one of classifying, clustering or indexing the subject genome sequence" are unclear as each of these would require different steps. The Office Action further states that the steps of "classifying, clustering or indexing" are not claimed, that the basis of "classifying, clustering or indexing" is not described and that the boundaries of each class, cluster of index are not defined. The Office Action further states that the limitation of "using uniform representation" is "not illuminating" as no specific method steps associated with such use are recited.

The Applicant respectfully disagrees. The present invention is not directed to new or specific algorithms for "classifying, clustering or indexing", *i.e.* defining classes, of sequences. Rather, the present invention is a method of analyzing a subject's genome sequence by assigning it to a class by ***using the "uniform representation***". As submitted above by the Applicant, the "uniform representation" refers to the transformation of genome sequences of variable lengths into algebraic constructs termed "feature vectors", each of which has the same number of elements. Not only does the specification describe how such transformation is achieved, but, in fact Claims 1, 10, 19 and 20 expressly recite the steps of such transformation.

Thus, for example, Claim 1 recites the step of:

> "comparing the respective representation of each known biological fragment from the set to a subject genome sequence, for each known biological fragment said comparing including
>
> > (i) counting the number of times the respective representation of the known biological fragment is found in the subject genome sequence; and
> >
> > (ii) from said counted number of times, forming a vector element, such that for each known biological fragment there is a respective vector element representing the number of times the respective representation of that known biological fragment is found in the subject genome sequence".

Following recitation of how to form vector elements, Claim 1 proceeds to recite how to form a feature vector:

> "from the formed vector elements, forming a vector having a length equal to the fixed number of known biological fragments in the provided set, such that the formed vector provides a uniform representation of the subject genome sequence".

Likewise, an apparatus Claim 10 also recites how a feature vector of a uniform representation is formed from the scores of the "known biological fragments":

> "the generated score is one of a probability of the subject genome sequence being generated by the known biological sequence or a counting of a number of occurrences of the known biological sequence found in the subject genome sequence".

Lastly, base method and apparatus claims also recite that the uniform representation is used as input for the step of analyzing or the element of analysis routine.

The Applicant submits that the statement in the Office Action that "the limitation of "using uniform representation" is "not illuminating" as no specific method steps associated with such use are recited" is without merit.

Nor can the statements that the steps of "classifying, clustering or indexing are not claimed", that "the basis of classifying, clustering or indexing" is not described" and that "the boundaries of each class, cluster of index are not defined" support rejection under 35 U.S.C. §112, second paragraph.

As described above, the present invention employs various classification techniques. In some embodiments, a ***known*** indexing technique is applied to a commercially available database. In other embodiments, ***known*** algorithms (PCA, k-mean clustering, nearest neighbor analysis or support vector machine) are used to create sequence classification. The specification as filed provides guidance as to the possible "classifying, clustering or indexing" algorithm that can be used to practice the present invention. It is well established in Patent Law that a specification need not include, and preferably omits, what is known to a skilled person. M.P.E.P. §2163 (II)(A)(2) specifies that

> "[i]nformation which is well known in the art need not be described in detail in the specification. See, e.g., Hybritech, Inc. v. Monoclonal Antibodies, Inc., 802 F.2d 1367, 1379-80, 231 USPQ 81, 90 (Fed. Cir. 1986)."

Written Description Guidelines Training Materials, available at,

http://www.uspto.gov/web/offices/pac/writtendesc.pdf,

released March 1, 2000, also specify in Example 8, page 35:

> "Where, as here (1) the inventive portion of the subject matter is disclosed and (2) any additional variability within the genus arises due to additional elements that are not part of the inventor's contribution, and when the level of knowledge and skill in the art would allow one skilled in the art to recognize that the applicant was in possession of the genus, the written description cannot be deemed defective."

Here, the inventive portion of the invention, the use of uniform representation, is disclosed. Any additional variability within the claimed methods arises out of additional elements, *i.e.* classification algorithms, that are not part of the inventor's contribution. The level of knowledge and skill in the art would allow one skilled in the art to practice the present invention, especially in view of the examples of the various classification techniques provided in the instant application.

The Applicant submits that one skilled in the art will appreciate that the metes and bounds of the present invention are defined by the ***use of the uniform representation*** for analyzing a subject's genome sequence, as described in the instant application, and not by the algorithms of "classifying, clustering or indexing". One skilled in the art will readily appreciate that any algorithm for "classifying, clustering or indexing" can be used to practice the present invention and therefore the steps of such algorithms need not be explicitly recited in the claims.

According to the foregoing, base method and apparatus Claims 1, 10, 19 and 20 and new Claims 34-36 are believed to distinctly claim subject matter which the Applicant believes to be his invention.

Reconsideration and withdrawal of the rejection is respectfully requested.

CONCLUSION

In view of the above amendments and remarks, it is believed that all now pending claims (Claims 1-5, 7-14, 16, 19-30 and 34-36) are in condition for allowance, and it is respectfully requested that the application be passed to issue. If the Examiner feels that a telephone conference would expedite prosecution of this case, the Examiner is invited to call the undersigned at (978) 341-0036.

Respectfully submitted,

HAMILTON, BROOK, SMITH & REYNOLDS, P.C.

By _____

Mary Lou Wakimura
Reg. No. 31,804
Tel. (978) 341-0036
Fax (978) 341-0136
Concord, MA 01742-9133

Dated: 8/20/04